ONLINE SUPPLEMENTAL MATERIAL

DETAILED METHODS

Study design and participants

The survey protocols, the instruments, and the process for obtaining informed consent were approved by the institutional review committees of the University of North Carolina at Chapel Hill and the China Center for Disease Control and Prevention.

Measurement of variables

For the food inventory, we measured all available foods in the household on a daily basis, and we considered wastage to estimate the total consumption. For the 24-hour recall, trained interviewers recorded the type, amount, type of meal, and place of consumption of all food items consumed. For dishes prepared at home, we estimated the amount for each person based on what he or she reported and the household food inventory.

We measured glucose in the serum with a glucose oxidase phenol 4-amnioantipyrine peroxidase kit (Randox, UK) in a Hitachi 7600 analyzer. We measured HbA1c in the whole blood by high-performance liquid chromatography with an automated glycohemoglobin analyzer (model HLC-723 G7, Tosoh, Tokyo, Japan). We measured insulin in the serum by radioimmunology in a gamma counter XH-6020 analyzer.

For physical activity, we collected detailed time spent and intensity levels for occupational and domestic activities and estimated metabolic equivalents per week.[1] We determined the level of urbanization with an urbanicity index that was developed for the CHNS. It includes components such as population density, economic activity, transportation infrastructure, sanitation, and housing types.[2] We estimated BMI from measured weight and height.

Statistical analysis

Dietary pattern

In a previous analysis conducted in the CHNS population, but in a different analytic sample as eligibility requirements were different, we derived a dietary pattern with reduced rank regression (RRR). Dieatary data was from 2006 and response variables from 2009.

The analysis included twenty-nine food groups. Because of the large proportion of nonconsumers, we categorized intake as binary: nonconsumers versus consumers for food groups with less than 80% of consumers and below versus above the median otherwise (rice, fresh nonleafy vegetables, and fresh leafy vegetables). With the residuals method, we adjusted the food groups by geographic region, urbanicity index, education, and income (we estimated residuals from a regression with the food group as the dependent variable and the demographic variables as the independent ones). We performed RRR on the residuals with PROC PLS and RRR option (SAS 9.3, SAS Institute Inc., Cary, North Carolina) with log-transformed HbA1c, HOMA-IR, and fasting glucose as response variables. As is customary when using RRR diet patterns, we retained only the first pattern, which explained most of the variation in the response variables (Batis et al., unpublished data, 2013).

For the present study we computed a score of this dietary pattern in our analytic sample from 1991 to 2006. To simplify the computation of the score and make it more interpretable over time, we did a weighted sum (using the loadings as weights) of the original binary food groups (instead of the residuals) (i.e., score = $\text{rice}_{(0,1)} * -0.21 + \text{wheat noodles}_{(0,1)} * 0.24 + \text{wheat flour}_{(0,1)} * 0.13 + \text{wheat buns and breads}_{(0,1)} * 0.36 \ldots$). We used the same loadings in all waves, so that we could assess changes on a score that represents the same dietary pattern across time.

Latent class trajectory analysis

For estimating each model, we specified 100 initial stage random starts, 10 final stage optimizations, and 10 initial stage iterations. We reran the five-class model with the second best seed value to confirm that the estimates were replicated and the solution was not local.

Sensitivity analysis

To address selection bias, we computed inverse probability weights (IPW), so that those included in the analysis represented all the eligible (included and excluded) participants.[3] Out of the 7,646 eligible participants, we included 7,506 that had complete information to compute the inverse probability weights. We fitted two logistic regressions, one to predict the probability of being selected and another to predict the probability of being selected conditional on the exposure and the covariates. Then we estimated stabilized weights as the ratio of these two probabilities. Because most of the excluded participants were excluded because they did not have exposure information (dietary pattern trajectories from 1991 to 2006), we used as exposure their dietary pattern score in 2009 to compute the IPW. Similarly, their covariate information for the IPW was from 2009. We reran the main analysis using the IPW and a robust variance estimator. We found that using the IPW did not change the interpretation of our findings. For example, the results for change in HbA1c for classes 2, 3, 4, and 5 compared to class 1 were -1.66 (95% CI = -3.45, 0.12), -0.89 (-2.08, 0.30), -2.44 (-3.68, -1.19), and -4.35 (-6.05, -2.65), respectively. This suggests that selection bias was minimal.

In another sensitivity analysis, we repeated the analysis including only subjects with four or more waves of dietary data (instead of three) (n = 3,383). We reran the latent class trajectory analysis for one to seven classes and found that the best fit was for five classes. The shapes of the trajectories were very similar to those of our main analysis. Also the association between the classes and the outcomes did not change importantly. For example, the results for change in HbA1c for classes 2, 3, 4, and 5 compared to class 1 were -1.66 (95% CI = -3.32, 0.01), -0.99 (-2.27, 0.28), -2.48 (-3.81, -1.15), and -5.26 (-7.16, -3.36), respectively.

REFERENCES

1. Ng SW, Norton EC, Popkin BM. Why have physical activity levels declined among Chinese adults? Findings from the 1991-2006 China Health and Nutrition Surveys. *Soc Sci Med.* Apr 2009;68(7):1305-1314.

2. Jones-Smith JC, Popkin BM. Understanding community context and adult health changes in China: Development of an urbanicity scale. *Social Science & Medicine.*71(8):1436-1446.

3. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* Sep 2004;15(5):615-625.

Supplemental Table 1. Dietary pattern derived from reduced rank regression in 2006

Food groups	Factor loadings*
Rice	**-0.22**
Wheat noodles	**0.30**
Wheat flour	0.08
Wheat buns and breads	**0.46**
Cakes, cookies, and pastries	0.00
Deep-fried wheat products	**0.22**
Corn and coarse grains	0.09
Starchy roots and tubers	0.11
Fresh legumes	**-0.24**
Dried legumes	0.05
Legume products	0.03
Nuts and seeds	-0.11
Starchy root products and tuber products	0.14
Fresh vegetables, nonleafy	0.01
Fresh vegetables, leafy	0.08
Pickled, salted, or canned vegetables	-0.01
Dried vegetables	0.19
Fruits	-0.08
High-fat red meat	0.02
Low-fat pork	0.18
High-fat pork	-0.08
Processed meats	-0.15
Organ meats	-0.01
Poultry and game	**-0.37**
Eggs	**-0.23**
Fish and seafood	**-0.29**
Soy milk	**0.24**
Animal-based milk	-0.14
Instant noodles and frozen dumplings	0.09

*Factor loadings > |0.20| are in bold.

Supplemental Table 2. Model fit and trajectory class characteristics

BIC	Sample-size adjusted BIC	Class	Proportion in each class	Intercept	Slope	Posterior probability	Entropy
23,606.67	23,581.25	1	1.00	-0.179	0.007	1.00	NA
19,938.50	19,903.55	1	0.41	0.059	0.014	0.90	0.72
		2	0.59	-0.347	0.001	0.93	
19,545.42	19,500.93	1	0.46	-0.176	0.006	0.77	0.60
		2	0.24	0.130	0.018	0.86	
		3	0.30	-0.432	-0.003	0.82	
19,483.25	19,429.24	1	0.17	0.154	0.020	0.82	0.61
		2	0.46	-0.336	0.001	0.79	
		3	0.33	-0.078	0.009	0.71	
		4	0.04	-0.582	-0.006	0.75	
19,444.58	19,381.03	1	0.18	0.151	0.021	0.78	0.57
		2	0.06	-0.341	0.036	0.56	
		3	0.24	0.026	-0.002	0.60	
		4	0.47	-0.323	-0.001	0.76	
		5	0.05	-0.592	-0.004	0.74	
19,449.75	19,376.67	1	0.04	-0.356	0.041	0.57	0.59
		2	0.01	-0.037	-0.041	0.63	
		3	0.06	-0.585	-0.004	0.72	
		4	0.18	0.155	0.021	0.78	
		5	0.46	-0.343	0.003	0.74	
		6	0.26	0.016	0.000	0.62	
19,457.84	19,375.23	1	0.05	-0.597	-0.004	0.72	0.58
		2	0.01	-0.703	0.075	0.68	
		3	0.45	-0.355	0.003	0.72	
		4	0.25	-0.157	0.016	0.56	
		5	0.01	-0.056	-0.037	0.62	
		6	0.19	0.150	0.022	0.78	
		7	0.05	0.136	-0.013	0.56	

BIC = Bayesian information criteria; NA = not applicable.

Supplemental Table 3. Association between dietary pattern score quartiles in 2006 and HbA1c, HOMA-IR, and prevalence of newly diagnosed diabetes in 2009

	HbA1c % change (95% CI)*	HOMA-IR % change (95% CI)*	Diabetes odds ratio (95% CI)
Dietary pattern score in 2006			
Quartile 1	0 (Ref.)	0 (Ref.)	1 (Ref.)
Quartile 2	2.07 (0.97, 3.16)	2.42 (-4.14, 8.97)	1.75 (1.04, 2.92)
Quartile 3	2.58 (1.61, 3.55)	5.04 (-1.95, 12.03)	2.11 (1.28, 3.48)
Quartile 4	3.92 (2.87, 4.97)	9.05 (1.79, 16.32)	2.02 (1.23, 3.32)

*Because HbA1c and HOMA-IR were natural log transformed, the regression coefficients were multiplied by 100 and interpreted as the percentage of change in the outcome for being in a given quartile compared to the reference quartile. Adjusted for age, gender, geographic region, income, urbanicity index, education, physical activity, smoking, and alcohol intake.

Supplemental Table 4. Independent association between relevant food groups and HbA1c

	% change in HbA1c (95% CI)[a]
Rice[b]	-2.64 (-4.03, -1.25)
Wheat noodles	3.06 (1.75, 4.37)
Wheat buns and breads	2.50 (0.91, 4.08)
Deep-fried wheat products	1.53 (-0.59, 3.66)
Fresh legumes	-0.64 (-1.98, 0.71)
Poultry and game	-1.95 (-3.86, -0.04)
Eggs and egg products	-0.53 (-2.04, 0.98)
Fish and seafood	-1.20 (-2.54, 0.14)
Soy milk	-0.85 (-3.84, 2.13)

[a]For a one-unit change in the proportion of participating waves (1991–2006) in which the food group was consumed; i.e., the average difference in percentage of change in HbA1c is 3.06 percentage points between those who had always consumed wheat noodles (proportion of consumption = 1) vs. those who never had (proportion of consumption = 0). We estimated each coefficient in a different model, adjusting by age in 2006, gender, geographic region, mean income, mean urbanicity index, mean physical activity, proportion of waves smoking, and proportion of waves with alcohol intake ≥ 3 times/week.
[b]Proportion below and above the median instead of proportion consumed.